ABSTRACT OF THE DETAILED DESCRIPTION

A computerized method is used to estimate the relative coverage of Web search engines. Each search engine maintains an index of words of pages located at specific URL addresses in a network. The method generates a random query. The random query is a logical combination of words found in a subset of the pages. The random query is submitted to a first search engine. In response a set of URLs of pages matching the query are received. Each URL identifies a page indexed by the first search engine that satisfies the random query. A particular URL identifying a sample page is randomly selected. A strong query corresponding to the sample page is generated, and the strong query is submitted to a second search engine. Result information received in response to the strong query is compared to determine if the second search engine has indexed the sample page, or a page substantially similar to the sample page. This procedure is repeated to gather statistical data which is used to estimate the relative sizes and amount of overlap of search engines.